A new glycoside hydrolase family reveals a novel catalytic apparatus and active site location for beta-propeller enzymes

Jose Munoz-Munoz[a]*, Alan Cartmell[a]*, Nicolas Terrapon[b] Bernard Henrissat[b,c,d], Harry J. Gilbert[a,1]

Running title: Novel rhamnosidase with atypical active site

Classification: BIOLOGICAL SCIENCES; Biochemistry

[a]*Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK;*

[b]*Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille University, F-13288 Marseille, France;*

[c]*INRA, USC1408 Architecture et Fonction des Macromolécules Biologiques, F-13288 Marseille, France,*

[d]*Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia;*

* Equal contribution

[1]Corresponding author. E-mail harry.gilbert@ncl.ac.uk

ABSTRACT

The human gut microbiota utilize complex carbohydrates as major nutrients. The requirement for an efficient glycan degrading systems exerts a major selection pressure on this microbial community. Thus, we propose that these bacteria represent a substantial resource for discovering novel carbohydrate active enzymes. To test this hypothesis we focused on enzymes that hydrolyse rhamnosidic bonds as cleavage of these linkages is chemically challenging and there is a paucity of information on L-rhamnosidases. Here we screened the activity of enzymes derived from the HGM bacterium *Bacteroides thetaiotaomicron*, which are upregulated in response to rhamnose-containing glycans. We identified a novel α-L-rhamnosidase, BT3686, which is the founding member of a new glycoside hydrolase family, GH-BT3686. In contrast to other rhamnosidases, BT3686 cleaved L-Rha-α1,4-D-GlcA linkages through a retaining double displacement mechanism. The crystal structure of BT3686 showed that the enzyme displayed a type A seven-bladed β-propeller fold. Mutagenesis and crystallographic studies, including the structure of BT3686 in complex with the reaction product GlcA, revealed a novel location for the active site among β-propeller enzymes cited on the posterior surface of the rhamnosidase. In contrast to the vast majority of glycoside hydrolases, the catalytic apparatus of BT3686 does not comprise a pair of carboxylic acid residues but, uniquely, a single histidine functions as the only discernable catalytic amino acid. Intriguingly the histidine, His48 is not invariant in GH-BT3686, however, when engineered into structural homologs lacking the imidazole residue, α-L-rhamnosidase activity was established. The potential contribution of His48 to the catalytic activity of BT3686 is discussed.

Key words: Human Gut Microbiota, L-rhamnosidase, Bacteroides thetaiotaomicron, Gum Arabic.

Significance Statement

The location of the active site of enzymes with the same fold is invariably conserved. The beta propeller fold exemplifies this feature with all functions located to what is termed their anterior surface. Here, however, we show that the active site of a novel glycoside hydrolase that adopts the beta propeller fold is located to the posterior surface of the α-L-rhamnosidase. The enzyme also displays a novel catalytic apparatus that utilizes a single histidine instead of the canonical pair of carboxylate residues deployed by the vast majority of glycoside hydrolases. The capacity to engineer catalytic functionality into the posterior surface of other family members provides insight into the evolution of this novel enzyme family.

"/body"

INTRODUCTION

The human large bowel is a highly competitive environment where limited carbon resources, in the form of complex glycans, are competed for by a vast number of microorganisms defined as the human gut microbiota or HGM (1). Genomic studies have shown that the HGM contain a large number of genes encoding carbohydrate active enzymes (CAZymes), with some organisms dedicating up to 20% of their genome to carbohydrate metabolism (2). Given that glycan metabolism exerts a major selection pressure on the HGM, this microbial ecosystem represents a substantial resource for discovering novel CAZymes.

The major CAZymes that depolymerize carbohydrate polymers are glycoside hydrolases (GHs), polysaccharide lyases (PLs) and lytic polysaccharide monooxygenases (LPMOs). These CAZymes are grouped into sequence-based families on the CAZY database (3). The structural fold, catalytic apparatus and mechanism are highly conserved within families. Substrate specificity can vary within a family, exemplified by family GH5, or be invariant as in family GH10 (3). GHs generally hydrolyse glycosidic bonds through acid-base assisted catalysis deploying either double or single displacement mechanisms leading to retention or inversion of anomeric configuration, respectively. In nearly all of the 135 GH families the catalytic apparatus comprises two carboxylate residues (4). Exceptions to the canonical catalytic apparatus, in addition to histidine that is described in depth in Results and Discussion, include enzymes in GH95 where asparagine residues function as the catalytic base (5), while the catalytic nucleophile in GH33, GH34 and GH83 sialidases and GH127 arabinofuranosidases are tyrosine (6) and cysteine, respectively (7).

L-Rhamnose (L-Rha*p*), is found in a wide range of glycans in plants particularly the rhamnogalacturonans (RGI and RGII) and arabinogalactan proteins (AGP), **Fig. S1** (8). AGP comprises a β–1,3-D-Gal backbone decorated with β–1,6-D-Gal side chains that, in turn, are substituted with a number of different sugars. AGPs

have been used for around 50 years as thickening agents in many common foods, one example being marshmallows. Thus strong selection pressures operate in the HGM to evolve enzymes that process these L-Rha*p*-containing glycans. The enzymatic hydrolysis of rhamnosidic bonds is particularly challenging. In line nucleophilic attack at C1 of rhamnose would involve steric clash with the axial O2 hydroxyl, analogous to the challenges in mannose chemistry (9). Distortion of the pyranoside ring, to place O2 pseudo equatorial, will be needed to minimise these clashes, as has been observed previously for an α-L-rhamnosidase from GH78 (10) and GH106 (11). Furthermore, understanding how microbes remove the rhamnose residues that modify the AGP side chains has significant implications in the food and biorefining industries and in understanding the ecology of the HGM. There is, however, a paucity of information on the enzymes that hydrolyse these linkages. Currently only three GH families, GH78, GH90 and GH106, contain enzymes that cleave α-L-rhamnosidic bonds (3), and in all three families the α-L-rhamnosidases display an inverting mechanism.

To explore whether organisms within the HGM have evolved novel mechanisms to hydrolyse rhamnosidic bonds we focused on the HGM bacterium *Bacteroides thetaiotaomicron*. When the organism was cultured on AGPs two genetic loci (termed polysaccharide utilization loci or PULs (12)) were upregulated (13). Here we identified a novel AGP-specific α-L-rhamnosidase, BT3686, encoded by one of the AGP-activated PULs, which is thus the founding member of a new GH family. The enzyme is specific for L-Rha*p*-α1,4-D-GlcA linkages present in AGP from Gum Arabic (defined henceforth as GA). BT3686 displays a seven-bladed β-propeller fold. Strikingly, the active site is in a novel location for β-propeller enzymes; the catalytic apparatus lacks the canonical acidic residues and the proposed catalytic histidine is not invariant within the family. A putative catalytic mechanism is proposed and, in concert with the analysis of other members of this new family, the evolutionary route leading to a catalytically competent α-L-rhamnosidase is discussed.

RESULTS AND DISCUSSION

Biochemical characterization of rhamnosidase activity: The genome of *B. thetatiotaomicron* contains two PULs that are upregulated by AGP (13). The loci contain several open reading frames of unknown function (*orfuk*), but lack genes encoding enzymes belonging to known L-rhamnosidase families (GH78, GH90 and GH106) even though L-Rha*p* is a prominent feature of several AGPs. For these reason we produced recombinant forms of the proteins encoded by the *orfuk*s and screened these potential enzymes for L-rhamnosidase activity using GA as the substrate. The data, **Fig. 1AB**, **Fig. 2SAB** and **Table 1**, showed that BT3686 released L-Rha*p* from GA. Thus, BT3686 is an exo-acting α-L-rhamnosidase, and is the founding member of a new GH family defined as GH-BT3686. BT3686 displays tight specificity for the leaving group as the enzyme was inactive against 4-nitrophenyl-α-L-Rha*p*, RGI, RGII or other α-L-Rha*p*-containing oligosaccharides, **Table S1**.

L-Rha*p* is found in several different contexts in GA; it can be linked α-1,2, α-1,4 and α-1,6 to galactose residues and α-1,4 to D-glucuronic acid (D-GlcA) (8). To determine the context of the rhamnosidic linkage targeted by BT3686, the backbone of GA was digested by an exo-acting β-1,3-galactanase that can accommodate Gal residues decorated at O6, and thus released the side chains from GA, which were purified. When BT3686 was incubated with the smallest oligosaccharide (disaccharide), the enzyme generated exclusively L-Rha*p* and D-GlcA in a 1:1 ratio, **Fig. 1B**. Kinetic analysis revealed BT3686 had a k_{cat}/K_m against the L-Rha-α–1,4-D-GlcA disaccharide that was ~10 fold lower than against GA, **Table 1**, alluding to the possibility of further positive subsites beyond the +1 subsite (in GHs the scissile bond is located between the -1 and +1 subsites). These data show that BT3686 is a α-L-1,4 rhamnosidase that has at

least two subsites, with rhamnose residing at -1 and D-GlcA at +1. BT3686 was not inhibited by EDTA and thus is not metal dependent, **Table 1,** and had a pH optimum of ~5.5, **Fig. S2C**. The pH curve contained only a descending limb with a pKa of ~6.0 suggesting that a single catalytic residue is involved consistent with the ionization of a histidine (see below).

NMR analysis of the rhamnosidase reaction: The catalytic mechanism of BT3686 was determined by ^{1}H ^{1}D NMR using GA as the substrate. The data, **Fig. 1C**, revealed the immediate generation of a signal with a chemical shift of 5.0 ppm, consistent with the α-anomer of L-Rha*p*. Within 30 min a second signal appeared with a chemical shift of 4.8 ppm corresponding to the β-anomer of L-Rha*p*, demonstrating that the sugar had undergone mutarotation (14). As L-Rha*p* is linked α-1,4 to D-GlcA, these data show that BT3686 mediates cleavage of the rhamnosidic linkage with retention of anomeric configuration, and thus displays a double displacement mechanism. This is in contrast to enzymes from the other three rhamnosidase families that mediate inversion of anomeric configuration upon bond cleavage ((15, 16) and A.S. Luis personnel communication).

Crystal structure of BT3686: BT3686 shared 86% identity with the *B. ovatus* protein BACOVA_03493 whose crystal structure was available (PDB 4IRT) but had no known function. Thus the crystal structure of BT3686 was solved by molecular replacement using BACOVA_03493 as the search model. BT3686 is a monomer comprising a single domain that adopts a seven bladed β-propeller fold, **Fig. 2A**. Each blade comprises four anti-parallel β-strands that extend out radially from a central core, giving the protein a spherical shape ~45 Å across and ~35 Å high. The propeller is closed by blade 7 in which the final strand is provided by N-terminal β-strand 1. The contribution of both the N- and C-termini to one of the blades in β-propeller proteins is termed molecular Velcro and stabilizes the fold, **Fig. 2** (17).

Position of the active site in this new family: GH-BT3686 comprises greater than 1000 proteins that show >30% sequence identity with the *B. thetaiotaomicron* enzyme. The amino acids that are invariant in the family are located in an extended pocket on the anterior surface of BT3686 in the centre of the propeller, **Fig. 3**, suggesting that this region comprises the active site. Indeed, the active site is located in the equivalent region on the anterior surface of eight "β-propeller" GH families, polysaccharide lyase family PL22 and even non-CAZyme β-propeller enzymes, **Fig 4** (18). Structural comparison of BT3686 with other beta propeller GHs reveals that highest amino acid conservation, including several invariant residues, is on this anterior side of the enzyme. This again points to an active site in the anterior pocket of these α-L-rhamnosidases, **Fig. 3**. Site directed mutagenesis, however, showed that none of the 19 amino acid substitutions completely ablated rhamnosidase activity, **Table S2**, a critical criterion for catalytic residues. These unexpected data suggest that the active site is not located on the anterior surface of the enzyme.

To identify the location of the active site extensive attempts were made to crystallize BT3686 in complex with L-Rha*p*, rhamnopyranose tetrazole (transition state mimic with a K_i of 10 μM) D-GlcA, and L-Rha*p*-α-1,4-D-GlcA. The only complex obtained was with D-GlcA, which should occupy the +1 subsite, **Fig. 2ABCD**. Surprisingly the uronic acid was bound in an extended shallow pocket/cavity in the centre of the ***posterior surface*** of the β-propeller, **Fig. 2B**. The D-GlcA makes multiple direct and solvent-mediated interactions with the enzyme, **Fig. 2C**. O1 hydrogen bonds via solvent to the backbone carbonyls of Asp203 and Glu204; O2 makes polar contacts with the backbone carbonyl of Asp 203 and the Nζ of Lys392; O3 interacts with the Oε1 of Glu389 and Nζ of Lys392, and the carboxylic acid group hydrogen bonds to the backbone amines of Asp151 and Gly152. The K392A mutation had no effect on catalytic activity whilst the E389A substitution caused a ~50 fold reduction in k_{cat}/K_m. The E389Q mutant, however, retained wild type activity demonstrating that Glu389 forms a productive hydrogen bond with O3. The perpendicular orientation of the

carboxylic acid group to the sugar ring, needed for coordination to the backbone amines of Asp151 and Gly152, may be facilitated by Water 250. O4 of the +1 D-GlcA points into the extended cavity, which may therefore comprise the active site (-1 subsite). Apolar interactions with the substrate may be mediated by Pro148 at the +1 subsite and likely Met95, and Val46 in the active site. Indeed the valine and methionine residues line a pocket that could accommodate the C6 methyl group of L-Rha*p*. There are a paucity of polar residues in the proposed -1 subsite. His48 and Glu389 are in the proposed active site, while Glu100, Asp151, Lys392 and Lys393 are in the vicinity of the putative catalytic centre. In addition to the analysis of the mutants K392A, E389A and E389Q, described above, the influence of H48A, H48Q, E100A, D151A, D151N and K393A substitutions on catalytic activity were also assessed, **Table S2.** Surprisingly the mutants E100A, D151N and K393A retained wild type activity, while the D151A mutation caused a ~six-fold reduction in k_{cat}/K_m. This suggests that a pair of carboxylate residues, a canonical feature of the catalytic apparatus of the vast majority of GHs (4), do not make a direct contribution to cleavage of rhamnosidic bonds by BT3686. The H48A and H48Q mutants of BT3686 were completely inactive, and CD spectra indicated that the amino acid substitutions did not alter the protein fold, **Fig. S2D**. These data support a direct role for His48 in catalysis. This is highly unexpected as histidine is rarely used as a catalytic residue in GHs (see below), and acid-base assisted hydrolysis of glycosidic bonds is generally mediated by two amino acids (4).

The Role of histidine: A BLAST analysis of the top 113 sequences reveals BT3686 to be highly conserved with the lowest identity being 59%. The apparent catalytic histidine, however, is only present in ~70% of sequences with the imidazole residue being replaced primarily by Gln (in ~25 % of sequences), and rarely by Asp, Phe, Tyr or Ser, **Fig. S3**. The lack of conservation of the proposed catalytic residue in a GH family, although not unprecedented, is extremely unusual; exceptions are found in GH1, GH23, GH43, where the lack of a catalytic residue influences specificity (19-21), and GH97 where Glu-Asp substitutions

alters mechanism (22). The lack of conservation of BT3686 His48 may indicate that not all the proteins in GH-BT3686 are catalytically active. To test this hypothesis we analysed the activity of three proteins in GH-BT3686, BACCELL_00856, BACINT_00347 and BACPLE_00338, in which the BT3686 catalytic histidine was replaced by Gln, Gln and Asp, respectively, but sequence identity with the *B. thetaiotaomicron* enzyme was >80%. We also evaluated the activity of BACOVA_03493 and HMPREF9455_02360, which contain a His equivalent to His48 and display 86% and 59% overall sequence identity with BT3686. Only BACOVA_03493 and HMPREF9455_02360 were shown to be functional α-L-rhamnosidases against GA, suggesting that a His equivalent to BT3686-His48 is required for enzymes to exhibit this activity, **Fig. S2A**. It is formally possible that other features of the three enzymes that lack the conserved histidine confer the observed lack of activity. To test this hypothesis the equivalent residue to His48 in BACCELL_00856, BACINT_00347 and BACPLE_00338 was replaced with histidine, generating the mutants Q48H, Q48H and D46H, respectively. All three mutants displayed α-L-rhamnosidase activity at rates similar to BT3686, BACOVA_03493 and HMPREF9455_02360, **Fig. S2B** and **Table 1**. This demonstrates that the wild type versions of these proteins were catalytically inactive only through the absence of the imidazole amino acid. To provide further support for the catalytic-histidine hypothesis the crystal structures of BACCELL_00856 and BACINT_00347 were solved and compared with BT3686 and BACOVA_03493. The root-mean square deviation of the four structures ranged from 0.3-0.9 Å, consistent with sequence identities >80%. In BT3686 the region that comprises the anterior extended pocket was disordered and thus could only be built partially. In contrast the equivalent region in the other three enzymes was highly ordered clarifying the topology of the anterior pocket. Although the anterior pockets displayed a high degree of conservation, no canonical GH catalytic apparatus was evident, **Fig. S4**, again indicting that this region does not house the active site. Apart from the replacement of His with Gln in BACCELL_00856 and BACINT_00347, the

proposed posterior active site cavity was conserved in the four enzymes, which is entirely consistent with the catalytic-histidine hypothesis.

As the anterior pocket is highly conserved in GH-BT3686, and is the location of the active site in other β-propeller GHs, it is likely that this region comprises the catalytic centre at least in progenitors of this family, and possibly in members of the current GH-BT3686 family, particularly those that do not display rhamnosidase activity. To evaluate the latter possibility BT3686 and a GH-BT3686 protein that lacked a histidine equivalent to His48 (BACINT_00347) were screened for enzyme activity against several AGPs and other plant cell wall polysaccharides, **Table S1**. The data revealed no activity against these glycans. Although this indicates that the anterior pocket of members of GH-BT3686 do not comprise a functional active site, it is formally possible that these proteins are catalytically competent, but the appropriate substrates were not evaluated here. Alternatively, it is possible that the progenitor of GH-BT3686 is indeed catalytically incompetent but displays some other function.

BT3686 is a glycoside hydrolase and not a phosphorylase: Histidine has been shown to fulfill a catalytic role in GH117 and GH3 enzymes. In the GH117 enzyme *Bp*GH117, an inverting exo-acting 3,6-anhydro-(1,3)-α-L-galactosidase, a histidine functions as the catalytic acid (23). In GH3 β-N-acetyglucosaminidases the catalytic acid-base glutamate, typical of enzymes within this family, is replaced with a histidine (24, 25). It was proposed that the imidazole residue fulfilled a catalytic acid-base function in a classical retaining hydrolase reaction (25), which was clearly demonstrated through NMR studies of the GH3 enzyme Hsero1941 (26). Compelling evidence by Macdonald and colleagues, however, showed that *Cellulomonas fimi* Nag3 (contains a histidine that functions as a catalytic acid-base) is a retaining phophorylases in which phosphorolysis cleaves the glycosyl-enzyme covalent intermediate (24). It was proposed that installing histidine as opposed to a carboxylate residue decreases the anionic environment of the active site allowing the entrance of phosphate. It

is formally possible, therefore, that BT3686 is also a retaining phosphorylase. The data presented here measured the hydrolase activity of the rhamnosidase and not the potential of the enzyme to act as a phosphorylase. Thus, the reaction was determined through oxidation of rhamnose at C1, which cannot occur when the sugar is phosphorylated. Furthermore, the NMR experiment, carried out in 20 mM sodium phosphate, showed that the rhamnose generated underwent mutarotation, which requires a non-phosphorylated anomeric carbon, and no signal corresponding to the H1 of α-rhamnose-1-phosphate (5.3 ppm) appeared, **Fig. 1C**. The activity of BT3686 against GA or a GA-derived trisaccharide did not alter when phosphate was titrated into these reactions, **Fig. S5AB**. Finally, BT3686 quantitatively released all the rhamnose present in a GA-derived trisaccharide in its non-phosphorylated form irrespective of the phosphate concentration of the reactions, **Fig. S5C**. Collectively, these data show that BT3686 is a GH and displays no phosphorylase activity.

Putative catalytic mechanism: The NMR data revealed that BT3686 performs catalysis with retention of the anomeric configuration, consistent with a double displacement retaining mechanism, **Fig. 1C**. This mechanism typically utilizes a pair of acidic amino acids, one acting as the acid/base and the other the catalytic nucleophile, which attacks the anomeric carbon of the sugar that participates in the scissile glycosidic bond. While His48 is the only candidate nucleophile, it is 7 Å from the O4 of GlcA (the oxygen involved in the scissile glycosidic linkage) and is thus too distant from the anomeric carbon of rhamnose to mount a direct nucleophilic attack. For such a reaction to occur requires movement of His48 and/or the substrate prior to catalysis. Although rare, retention of the anomeric configuration can also be obtained via an epoxide mediated mechanism, **Fig. S6B**, as predicted for GH99 endo-α1,2-mannosidases (27), which may be more consistent with the position of the catalytic histidine. This mechanism is only possible on sugars where the O2 and the leaving group are situated in an anti-periplanar configuration as is the case for α-L-rhamnose.

Whether His48 functions as the catalytic nucleophile in a standard double displacement mechanism or in the highly unusual epoxide mechanism, a catalytic acid/base residue is required to protonate the leaving group and to activate a water molecule in the second phase of the reaction. Mutagenesis studies indicate that BT3686 contains no such candidate residue. The carboxylate of D-GlcA may fulfill this function by protonating its own O4. Upon glycosidic bond cleavage the D-GlcA would need to be retained in the +1 subsite to contribute to deglycosylation by activating an incoming water molecule. In this respect it is interesting to note the proposal that the leaving group (cellobiose) of a retaining GH7 cellobiohydrolase contributes to the positioning of the catalytic water prior to departure from the active site (28)

Phylogeny: An alignment-based phylogenetic tree, **Fig. S7**, shows the presence of BT3686 homologs in several bacterial phyla (Verrumicrobia, Proteobacteria and Actinobacteria) and in fungi (both basidiomycetes and ascomycetes). The strong conservation of the anterior pocket opposite to the active site contrasts sharply with the apparent relaxed evolutionary constraints that apply to the rhamnosidase active site of BT3686. Importantly all eukaryotic/bacterial phyla display a variety of residues that align to His48 of BT3686, and rarely a histidine, suggesting that the ancestral function of the family was different from that of BT3686 (and probably still is for most members of the family). The phylogenetic tree, **Fig. S7**, shows that His48 is rather well conserved across Bacteroides, suggesting that the particular active site of BT3686 is a recent innovation (neofunctionalization). The recent age of this event is compatible with the functional switch after mutation from Q to H in either *B. intestinalis* or *B. cellulosilyticus*. What is unclear at present is whether the unknown ancestral function is conserved or lost in the Bacteroides clade. Finally, the genetic context of genes encoding BT3686 homologs are broadly similar irrespective of the rhamnosidase activity of the cognate enzyme. The genes are located in loci that appear to encode arabinogalactan degrading systems exemplified by enzymes belonging to families GH43, GH2, GH16 and GH105.

Conclusions: This study reveals a new α-L-rhamnosidase family that targets L-Rha*p*-α-1,4-D-GlcA linkages. GH-BT3686 has significantly diverged from typical GHs. Unexpectedly the active site appears to be in a unique location for β-propeller GHs and, indeed, all enzymes that display this fold. Furthermore, the retaining enzyme family does not appear to contain a canonical catalytic apparatus comprising two carboxylate residues. In contrast, the catalytic apparatus of GH-BT3686 contains a histidine. Although this amino acid has been shown to be a key catalytic residue in some PL families, there is no precedent for the imidazole residue contributing to the catalytic apparatus of natural GHs. A further surprising feature of the GH-BT3686 family is the significant number of proteins that lack the catalytic histidine and appear to be inactive, and it is particularly intriguing that catalytic competence can be installed by introducing the catalytic histidine.

AGPs are important polysaccharides in plant biology and in the food industry. Their structures, however, are very diverse. The novel activity displayed by BT3686 and other family members makes an important contribution to the toolbox of biocatalysts available to dissect the structure of these complex glycans and to generate industrially relevant bespoke AGP-derived oligosaccharides. The structural and biochemical data reported here provides a framework for investigating the full range of activities displayed by members of this novel GH family.

Materials and Methods

Protein production and X-ray crystallography: Proteins were produced in *Escherichia coli* using standard expression vectors and IMAC purification methods. Proteins were crystallized in sitting drop plates. Structure determination was by molecular replacement. Structure statistics and proten databank accession codes are in **Table S3** and detailed methods are in ***SI Appendix.***

Enzyme assays: Activity assays used GA or GA-derived oligosaccharides as substrates. The reaction products were identified by TLC or HPAEC. The activity of the enzymes was determined using the L-Rhamnose detection kit from Megazyme International. Briefly L-rhamnose dehydrogenase, which was included in the assays oxidised L-rhamnose to L-rhamno-1,4-lactone with concomitant generation of NADH, which was assay at 340 nm. The rhamnosidase reactions were also monitored by NMR. Detailed methods can be found in ***SI Appendix***.

Acknowledgments

This work was supported by a grant to all the authors [European Union's Seventh Framework Program (FP/2007/2013)/European Research Council (ERC) Grant Agreement 322820].

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author's contributions

JLM-M, biochemistry of the rhamnosidase; AC, the crystal structure of the enzyme. NT and BH performed the phylogeny and HJG supervised the work of JLM-M and AC and contributed to writing the manuscript.

References

1. Arumugam M*, et al.* (2011) Enterotypes of the human gut microbiome. *Nature* 473(7346):174-180.
2. Flint HJ, Scott KP, Duncan SH, Louis P, & Forano E (2012) Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3(4):289-306.
3. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42(Database issue):D490-495.

4. Zechel DL & Withers SG (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* 33(1):11-18.
5. Nagae M, *et al.* (2007) Structural basis of the catalytic reaction mechanism of novel 1,2-alpha-L-fucosidase from Bifidobacterium bifidum. *J Biol Chem* 282(25):18497-18509.
6. Vocadlo DJ & Davies GJ (2008) Mechanistic insights into glycosidase chemistry. *Curr Opin Chem Biol* 12(5):539-555.
7. Ito T, *et al.* (2014) Crystal structure of glycoside hydrolase family 127 beta-l-arabinofuranosidase from Bifidobacterium longum. *Biochem Biophys Res Commun* 447(1):32-37.
8. Nie S-P, *et al.* (
9. Speciale G, Thompson AJ, Davies GJ, & Williams SJ (2014) Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr Opin Struct Biol* 28:1-13.
10. Fujimoto Z, *et al.* (2013) The structure of a Streptomyces avermitilis alpha-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J Biol Chem* 288(17):12376-12385.
11. Ndeh D, *et al.* (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions *Nature*:*in the press*.
12. Martens EC, Koropatkin NM, Smith TJ, & Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* 284(37):24673-24677.
13. Martens EC, *et al.* (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* 9(12):e1001221.
14. De Bruyn A & Anteunis M (1976) 1H-N.m.r. study of L-rhamnose, methyl alpha-L-rhamnopyranoside, and 4-o-beta-D-galactopranosyl-L-rhamnose in deuterium oxide. *Carbohydr Res* 47(1):158-163.
15. Pitson SM, Mutter M, van den Broek LA, Voragen AG, & Beldman G (1998) Stereochemical course of hydrolysis catalysed by alpha-L-rhamnosyl and alpha-D-galacturonosyl hydrolases from Aspergillus aculeatus. *Biochem Biophys Res Commun* 242(3):552-559.
16. Steinbacher S, *et al.* (1996) Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors. *Proc Natl Acad Sci U S A* 93(20):10584-10588.
17. Neer EJ & Smith TF (1996) G protein heterodimers: new structures propel new questions. *Cell* 84(2):175-178.
18. Wintjens R, *et al.* (2006) Crystal structure of papaya glutaminyl cyclase, an archetype for plant and bacterial glutaminyl cyclases. *J Mol Biol* 357(2):457-470.
19. Artola-Recolons C, *et al.* (2014) Structure and cell wall cleavage by modular lytic transglycosylase MltC of Escherichia coli. *ACS Chem Biol* 9(9):2058-2066.
20. Botti MG, Taylor MG, & Botting NP (1995) Studies on the mechanism of myrosinase. Investigation of the effect of glycosyl acceptors on enzyme activity. *J Biol Chem* 270(35):20530-20535.

21. Jiang D, *et al.* (2012) Crystal structure of 1,3Gal43A, an exo-beta-1,3-galactanase from Clostridium thermocellum. *J Struct Biol* 180(3):447-457.
22. Gloster TM, Turkenburg JP, Potts JR, Henrissat B, & Davies GJ (2008) Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chem Biol* 15(10):1058-1067.
23. Hehemann JH, Smyth L, Yadav A, Vocadlo DJ, & Boraston AB (2012) Analysis of keystone enzyme in Agar hydrolysis provides insight into the degradation (of a polysaccharide from) red seaweeds. *J Biol Chem* 287(17):13985-13995.
24. Macdonald SS, Blaukopf M, & Withers SG (2015) N-acetylglucosaminidases from CAZy family GH3 are really glycoside phosphorylases, thereby explaining their use of histidine as an acid/base catalyst in place of glutamic acid. *J Biol Chem* 290(8):4887-4895.
25. Litzinger S, *et al.* (2010) Structural and kinetic analysis of Bacillus subtilis N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *J Biol Chem* 285(46):35675-35684.
26. Ducatti DR, Carroll MA, & Jakeman DL (2016) On the phosphorylase activity of GH3 enzymes: A beta-N-acetylglucosaminidase from Herbaspirillum seropedicae SmR1 and a glucosidase from Saccharopolyspora erythraea. *Carbohydr Res* 435:106-112.
27. Thompson AJ, *et al.* (2012) Structural and mechanistic insight into N-glycan processing by endo-alpha-mannosidase. *Proc Natl Acad Sci U S A* 109(3):781-786.
28. Knott BC, *et al.* (2014) The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J Am Chem Soc* 136(1):321-329.
29. Rogowski A, *et al.* (2014) Evidence that GH115 alpha-glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility. *J Biol Chem* 289(1):53-64.
30. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125-132.
31. Anonymous (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760-763.
32. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
33. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
34. Murshudov GN, *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):355-367.
35. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.
36. Chen VB, *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12-21.

Table 1. Activity of BT3686 and homologs with various substrates.

Enzyme	Substrate	k_{cat}/K_m (min^{-1} M^{-1})
BT3686	GA (pH 5.5)	$5.77 \times 10^3 \pm 4.23\ 10^2$
BT3686 + 10 mM EDTA	GA (5.5)	$5.65 \times 10^3 \pm 1.14\ 10^2$
BT3686	GA (pH 7.0)	$7.80 \times 10^2 \pm 5.6 \times 10^1$
BT3686	Disaccharide	$1.36 \times 10^1 \pm 0.7\ 10^0$
BT3686	Trisaccharide	$5.84 \times 10^1 \pm 1.4\ 10^0$
BT3686	Tetrasaccharide	$1.97 \times 10^2 \pm 7.5\ 10^0$
BT3686	Heptasachharide	$2.5 \times 10^1 \pm 2.1\ 10^0$
BACOVA_03493	GA	$1.68\ 10^3 \pm 9.80\ 10^1$
BACCELL_00856	GA	n/a
BACINT_00347	GA	n/a
BACPLE_00338	GA	n/a
HMPREF9455_02360	GA	$1.40 \times 10^3 \pm 1.75 \times 10^2$
BT3686 H48A	GA	n/a
BACOVA_03493 H48A	GA	n/a
BACCELL_00856 Q48H	GA	$9.58\ 10^2 \pm 9.90 \times 10^1$
BACINT_00347 Q48H	GA	$1.20 \times 10^3 \pm 1.32 \times 10^2$
BACPLE_00338 D46H	GA	$1.57 \times 10^3 \pm 1.45 \times 10^2$
HMPREF9455_02360 H46A	GA	n/a

GA; Gum Arabic, Disaccharide (Rhamnose-α-1,4-GlcA), Trisaccharide (Rhamnose-α-1,4-GlcA-β-1,6-Gal), Tetrasaccharide (Rhamnose-α-1,4-GlcA-β-1,6-Gal-β-1,6-Gal), Heptasaccharide (Rhamnose-α-1,4-GlcA-(α-Ara_3)-β-1,6-Gal-β-1,6-Gal). All reactions were carried out in 20 mM Sodium phosphate pH 7.0 and 150 mM NaCl. n/a: No activity detected.

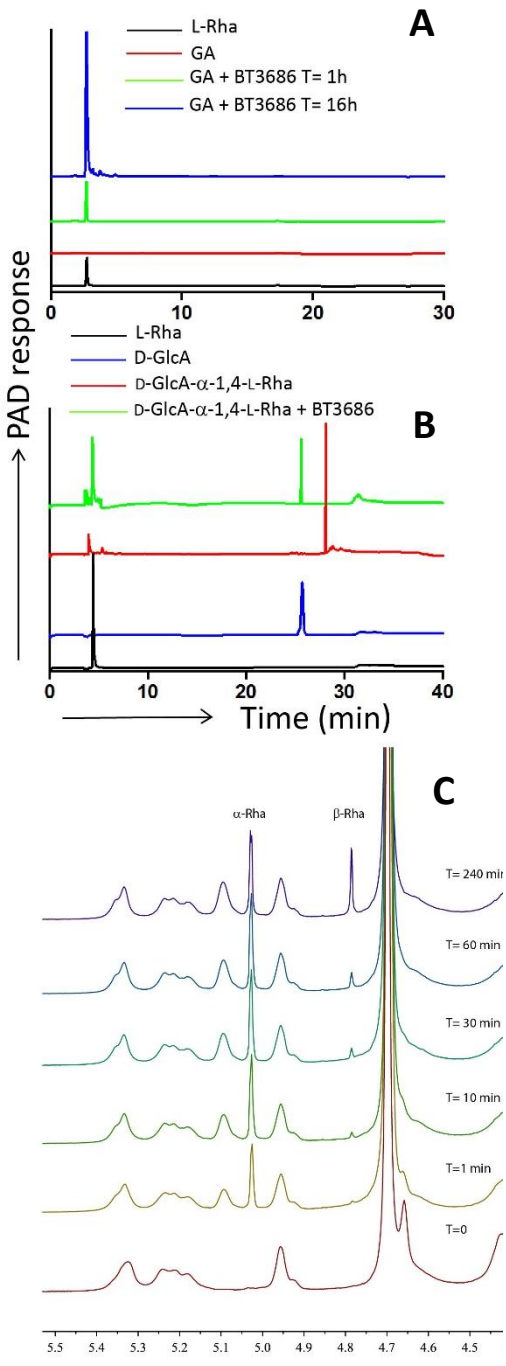
Figure Legends

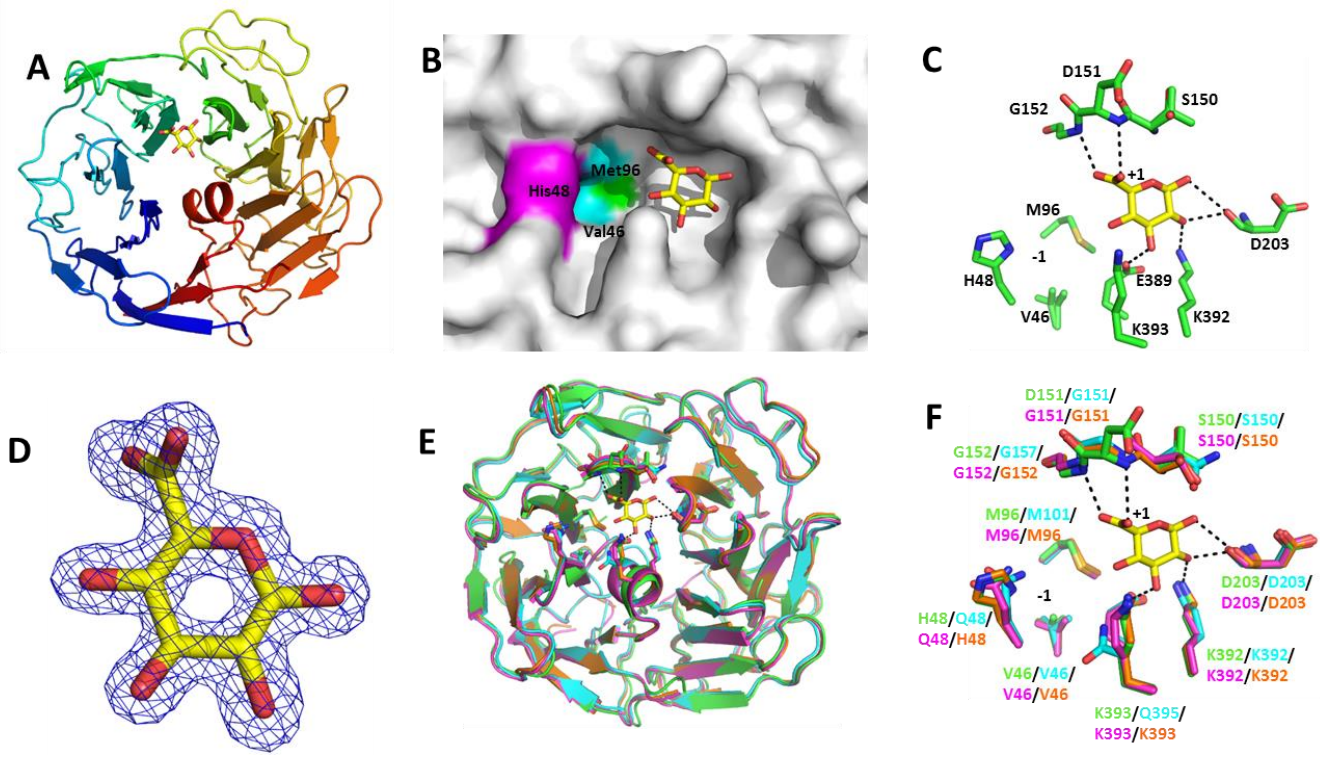
Fig. 1. Biochemical properties of BT3686. ***Panel A***. Time course of GA treated with BT3686, analyzed by HPAEC. ***Panel B***. Cleavage of the disaccharide Rha-α-1,4-GlcA by BT3686. ***Panel C***. H^+-NMR spectra of GA treated with BT3686 in D_2O. All reactions were carried out in 20 mM sodium phosphate pH 7.0 supplemented with 150 mM NaCl. Enzyme concentration used was 1 μM for ***Panel A*** and ***Panel B*** and 30 μM for ***Panel C***. For ***Panel A*** and ***Panel C*** 10 mM GA was used whilst 40 μM of the disaccharide was digested in ***Panel B***.

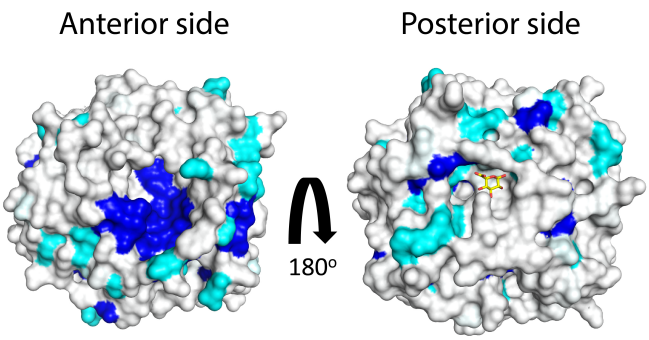
Fig. 2. The crystal structure of BT3686, BACINT_00347, BACCELL_00856 and BACOVA_03493. ***Panel A***. Schematic of BT3686 revealing the seven-bladed propeller fold with the colour ramped from *blue* at the N terminus to *red* at the C terminus. ***Panel B***. Solvent exposed surface representation of BT3686 with glucuronic acid bound in the +1 subsite. The critical histidine is coloured magenta and the hydrophobic residues at the bottom of the active site are in green (methionine) and cyan (valine). ***Panel C***. 3D location of the polar residues in BT3686 (coloured green) that interact with glucuronic acid or are in the vicinity of the active site. ***Panel D***. Electron density map ($2F_o - F_c$) of the glucuronic acid contoured at 1.0 σ (0.37 *e*/Å^3). ***Panel E*** and ***Panel F*** show an overlay of the fold (***Panel E***) and 3D location of active site residues (***Panel F***) of BT3686, BACINT_00347 BACCELL_00856 and BACOVA_03493 with carbons colored green, cyan, magenta and orange, respectively.

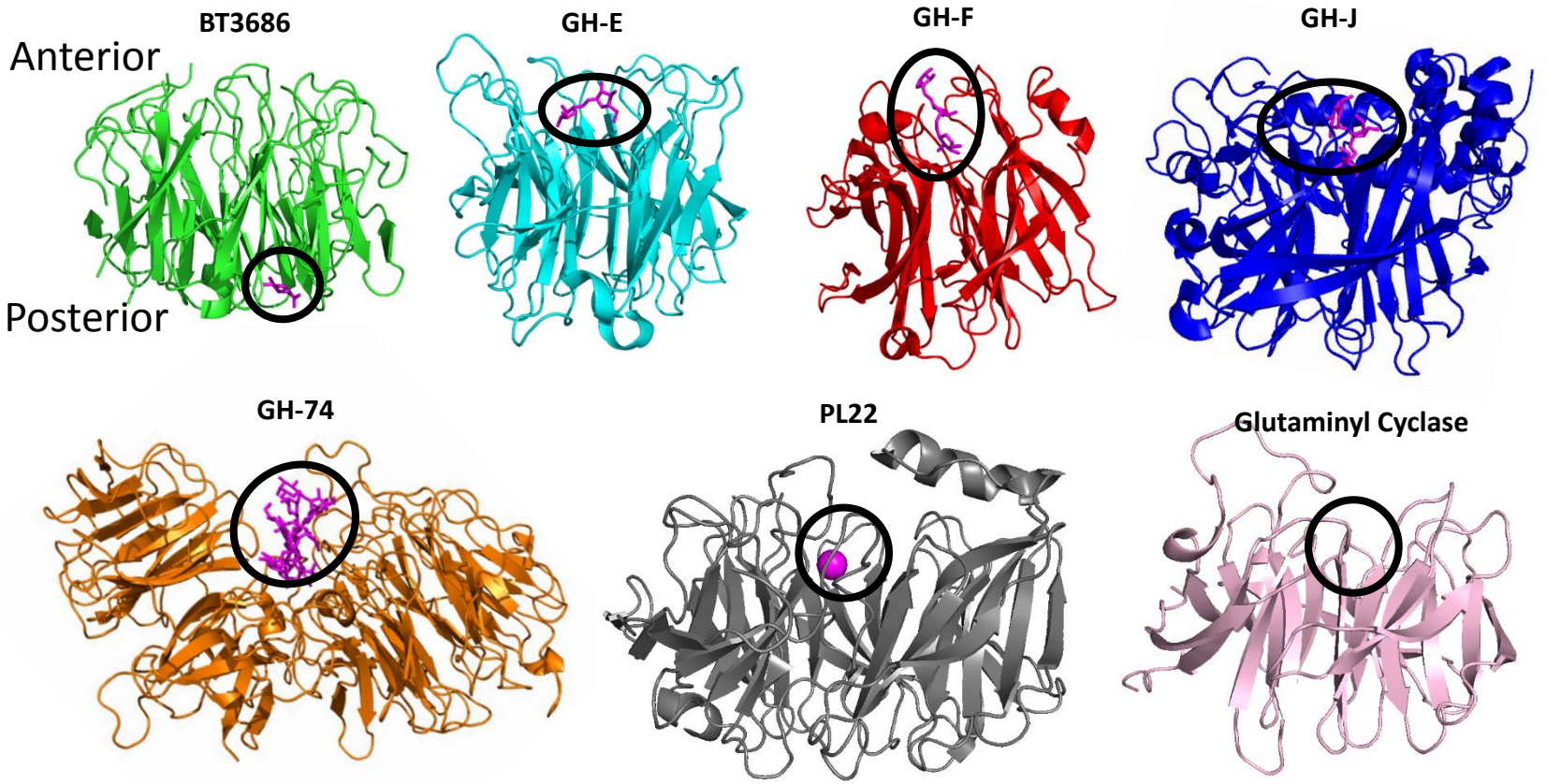
Fig. 3. Surface representation of sequence conservation at the anterior and posterior surfaces of BACOVA_03493. Residues were coloured relative to their conservation within the GH-BT3686 family. Dark blue signifies amino acids that are invariant and cyan identifies residues that are 40-60% conserved. The GlcA bound to the +1 subsite is depicted in stick format with carbons coloured yellow.

Fig. 4. Diverse examples of known enzymes with beta-propeller folds demonstrating the conservation of the active site location. BT3686 is shown in green; representatives of clan GH-E (GH93 3A72), GH-F (GH43 3AKH), clan GH-J (GH68 3BYN), GH74 (5FKS), PL22 (3PE7) and gutaminyl cyclases (3MBR) are shown in cyan, red, blue, orange, black and pink, respectively. All ligands are magenta and are highlighted with black circles to show the active site location. All proteins are orientated the same way.


A
L-Rha
GA
GA + BT3686 T= 1h
GA + BT3686 T= 16h
0
10
20
30
PAD response
L-Rha
D-GlcA
D-GlcA-α-1,4-L-Rha
D-GlcA-α-1,4-L-Rha + BT3686
B
0
10
20
30
40
Time (min)
C
α-Rha
β-Rha
T= 240 min
T= 60 min
T= 30 min
T= 10 min
T=1 min
T=0
5.5
5.4
5.3
5.2
5.1
5.0
4.9
4.8
4.7
4.6
4.5


A
B
His48
Met96
Val46
C
D151
S150
G152
+1
M96
D203
-1
H48
E389
V46
K393
K392
D
E
F
D151/G151/
G151/G151
S150/S150/
S150/S150
G152/G157/
G152/G152
M96/M101/
M96/M96
+1
D203/D203/
D203/D203
-1
H48/Q48/
Q48/H48
K392/K392/
K392/K392
V46/V46/
V46/V46
K393/Q395/
K393/K393


Anterior side
Posterior side
180°


BT3686
Anterior
Posterior
GH-E
GH-F
GH-J
GH-74
PL22
Glutaminyl Cyclase